Cry Toxins and Proteinase Inhibitors in Transgenic Plants Do Have Non-Zero Effects on Natural Enemies in the Laboratory: Rebuttal to Shelton et al. 2009

DAVID A. ANDOW,[1] GABOR L. LÖVEI,[2] AND SALVATORE ARPAIA[3]

Environ. Entomol. 38(6): 1528–1532 (2009)

A main point of our recent paper (Lövei et al. 2009) is that there are non-neutral effects of Cry toxins and proteinase inhibitors (PIs) on natural enemies in the laboratory and that the pattern of responses is complex and needs additional analysis. Shelton et al. (2009) aggressively attacked this conclusion. They claimed that all negative effects of Cry toxins are caused by effects of sublethally affected hosts and prey. We suggested in Lövei et al. (2009) and reiterate here that the actual situation is not that simple when laboratory studies are considered. We made our point by using statistical meta-analysis to show that there are more nonzero effects of Cry toxins and PIs on natural enemies than expected under a statistical null hypothesis that all observed effects were zero. The interested reader may want to examine the longer history of some of these issues (Lövei and Arpaia 2005; Andow et al. 2006; Romeis et al. 2006a, b).

In our rebuttal, we first address the deeper, fundamental questions raised by Shelton et al. (2009) about the value of meta-analysis and then proceed to rebut the core criticisms about our statistical methods. Although Shelton et al. (2009) raised many other issues, we limited our rebuttal to these central issues; our lack of comment does not imply agreement with their other complaints.

Strengths of Meta-Analysis

Shelton et al. (2009) make two criticisms of our work that are, in actuality, more fundamental criticisms of meta-analysis. These criticisms are made, in part, to defend the methods used and conclusions reached in reviews by O'Callaghan et al. (2005) and Romeis et al. (2006b), neither of which are based on meta-analyses. First they argued that nonsignificant *P* values are "devoid of futher meaning and interpretation" (Shelton et al. 2009, p. 318), and second, they argue that the author's conclusions should be given greater standing than a data-driven reading of the quantitative data (Shelton et al. 2009, p. 318). Both of these are fundamental criticisms of meta-analysis, so we address them first.

One of the most confusing aspects of meta-analysis is how it can take several statistically nonsignificant results and find statistical significance. It would seem that if 10 laboratories performed the same study and each laboratory found statistically nonsignificant results, the evidence for nonsignificance should be overwhelming. Such reasoning ignores the accumulation of sample size. Suppose instead, that one laboratory had performed the same experiment 10 times, each time finding nonsignificant results. If all 10 experiments had *P* values between 0.1 and 0.3, pooling the data could give rise to a statistically significant result because the pooled data have 10 times the sample size of each individual experiment. In a similar way, meta-analysis is a method for pooling the results from several laboratories to see if the combined results are significant even when none of the individuals studies was significant. Specifically, if the 10 laboratories produced nonsignificant *P* values uniformly distributed between 0.1 and 0.3, the combined data would have an expected *P* value of 0.00036, which is highly significant. Meta-analysis does just what Shelton et al. (2009) said we should not do—combine several nonsignificant studies to find possible statistical significance.

Meta-analysis also provides an objective way to combine the results from multiple authors because it relies on the data and not on the interpretation of the author (Romeis et al. 2006b) and avoids interpretive pitfalls. Thus, although Shelton et al. (2009) agreed with the summary statement of Bai et al. (2005) that "*Bt* rice pollen had no negative impacts on *P. japonica* fitness," such a statement may not hold up when the data are combined with those of others and are compared objectively using meta-analytic methods. Thus, meta-analysis rejects the notion that statistical evidence is found in the inferences and conclusions by an author about his/her own data; instead, evidence is found in the data published by that author. This point was not accepted by Shelton et al. (2009), but all scientists know that, whereas the interpretations may be the most interesting part of a paper, the data are primary.

[1] Department of Entomology and Center for Community Genetics, University of Minnesota, St. Paul, MN.

[2] Department of Integrated Pest Management, Aarhus University, Faculty of Agricultural Sciences, Flakkebjerg Research Centre, Slagelse, Denmark.

[3] ENEA National Agency for New Technologies, Energy and Environment, Research Centre Trisaia, Rotondella, Italy.

0046-225X/09/1528–1532$04.00/0 © 2009 Entomological Society of America

Modified Statistical Inference for Meta-Analysis

The core criticism of Shelton et al. (2009) concerns our statistical methods. If our statistical methods are valid, as we shall show in the following, the evidence from the existing laboratory studies implies that "Cry toxins and proteinase inhibitors often have non-neutral effects on natural enemies [in laboratory studies]" (Lövei et al. 2009, p. 293). None of the remaining issues raised by Shelton et al. (2009) would erase this important fact. Under "Prey/host-quality mediated effects," Shelton et al. (2009) argued (1) that study design is important for understanding the direct effects of Cry toxins on natural enemies and (2) that our meta-analysis should have distinguished bitrophic and tritrophic effects on natural enemies (see below). Under "Ecological relevance and risk assessment," they also argued that (3) results from laboratory studies do not necessarily imply that there will be risks in the field (our paper makes no contradictory statement) and (4) that our meta-analysis should have distinguished among kinds of proteinase inhibitors (this is desirable, but the data do not allow at present an analysis finer than Lövei et al. 2009, Table 6). Even if they were correct, it is extremely important to note that none of these additional complaints would contradict our finding of non-neutral effects of Cry toxins and PIs on natural enemies in the laboratory, and a more nuanced discussion becomes necessary.

Measure of Scaled Effect Size. In our paper, we stated that our effect size measure is "similar" to Hedges' *g* (Lövei et al. 2009, p. 293) to foreshadow that we would not be conducting a classic meta-analysis. Shelton et al. (2009) misinterpreted us; they added the words "but not the same as" (in parentheses, their p. 318), which is not in our original text. By doing this, they seek to cast doubt on the validity of our statistical methods.

Actually, our basic effect size statistic is mathematically equivalent to Hedges' *g*. Scaled Hedges' *g* is defined as

$$g = \sqrt{n}(\bar{x}_t - \bar{x}_c)/sd_p, \qquad [1]$$

where $\bar{x}_t$ is the sample mean for the treatment group, $\bar{x}_c$ is the sample mean for the control group, SD_p is the pooled SD of the two samples, and n is the sum of the number of replicate observations in the treatment and control (for simplicity, we show the case $n_t = n_c$) (Hedges and Olkin 1985). Our statistic is

$$(\bar{x}_t - \bar{x}_x)/SE_p, \qquad [2]$$

where SE_p is the pooled SE. This can be seen from p. 294 (Lövei et al. 2009), where we explain in detail how we handled the treatment and control means (giving a measure of effect size), and in our p. 295, where we state that the effect size is divided by SE_p to classify the reponse. Basic statistical theory defines

$$SE_p = sd_p/\sqrt{n}. \qquad [3]$$

When equation (3) is substituted into equation (2), we recover the right-hand side of equation (1), proving that our statistic is equal to scaled Hedges' *g*.

Hedges' *g* and Hedges' *d*. Marvier et al. (2007), Wolfenbarger et al. (2008), Duan et al. (2008), Nguyên et al. (2008), and Naranjo (2009) used Hedges' *d* as a measure of scaled effect size. Hedges' *d* is equal to Hedges' *g* multiplied by a correction term for small sample sizes. Specifically,

$d = Jg$, where

$$J = 1 - \frac{3}{4(n-2)-1}$$

is the correction term, with n defined as above (Hedges and Olkin 1985). When n is small (<15), the correction term J is essential, and when n is large (>20), J is negligible ($0.96 < J < 1$). Small n are typical in field experiments (Marvier et al. 2007, Duan et al. 2008, Nguyên et al. 2008, Wolfenbarger et al. 2008, Naranjo 2009) and some laboratory experiments on honey bees (Duan et al. 2008). Meta-analysis also requires that the scaled effect size is approximately normal. Both Hedges' *g* and Hedges' *d* have noncentral *t*-distributions. Under the null hypothesis of zero effect size (no difference between the treatment and the control), both are approximately normal when n is large (>20). For laboratory studies on natural enemies, n typically exceeded 20 (note that this would have only 10 replicate observations in the treatment and 10 in the control). Thus, not using the correction term does not compromise our analysis.

Statistical Inference. Because Hedges' *g* is approximately normal for large sample sizes under the null hypothesis of zero effect size for all observations, we posed the statistical question: does the distribution of observed effect sizes follow a normal distribution? If they do not follow a normal distribution and there are fewer than expected nonzero effects (treatment is different from control), we may conclude that there are nonzero effects in the data. An examination of the deviations from normality allows us to infer how often and in what ways the treatment differs from the control. We tested this null hypothesis by binning the effect sizes into five response groups based on SD units of the standardized normal distribution (Lövei et al. 2009, p. 295). This mathematical operation is conceptually identical to using individual effect sizes to calculate a mean effect size; we ascribed the same importance to the individual effect sizes as is done in classical meta-analysis. This operation allows us to calculate expected values for each bin and to convert the problem into a (admittedly not very powerful, see p. 298 in Lövei et al., 2009) contingency table analysis. Thus, there is nothing flawed in our analysis. Indeed our analysis is statistically conservative and has the additional property of being insensitive to extreme values. Thus, we showed that the existing studies on Cry toxins and PIs have more than expected nonzero effects on natural enemies in laboratory studies.

Still, we appreciate that a reader may wonder how is it that two meta-analytic views (ours and Naranjo 2009) of the almost identical corpus of work give divergent results. The classical meta-analysis approach used by Marvier et al. (2007), Wolfenbarger et al. (2008), Nguyên et al. (2008), and Naranjo (2009) summarized responses using an average effect size, which can be a flaw when aggregating diverse responses of multiple species studied with a variety of experimental designs, which all of these published meta-analyses do. A focus on the average can lead to faulty inference. Consider a hypothetical example. Suppose we know that the real effect of a certain chemical is to cause the egg shells of peregrine falcons to become 25% thinner and that this chemical also causes the egg shells of dodos to become 25% thicker. By using the average, Shelton et al. (2009) would have us take the mean (=0% change in egg shell thickness) and declare the toxin safe. Our approach highlights the real variation among responses. This variation is the *raison d'être* of risk assessment.

Nonindependence of Measures of Effect Size

Shelton et al. (2009) also argued that the lack of statistical independence in the data biases our analysis to yield false positives (type I error). The issue of independence is a vexed problem in ecological meta-analysis. There are many kinds of statistical dependence that remain uncontrolled in nearly all ecological meta-analyses, and these problems are evident in all published meta-analyses of the effects of genetically engineered crops on nontarget species (Marvier et al. 2007, Nguyên et al. 2008, Wolfenbarger et al. 2008, Naranjo 2009), except for Duan et al. (2008). For example, meta-analyses of field experiments treat the different species density responses in the same experiment as if they were independent, despite the fact that we know that species interact and can strongly affect each other's densities. Results from experiments conducted in the same environmental chamber are also likely to be correlated. Arguably, even results from the same primary investigator are likely to be correlated because of individualistic variation in technique and materials. Nearly all ecologically based meta-analyses have significant nonindependence among the observations. Meta-analysis theory has not addressed nonindependence, so there has been no way to address it except (unsatisfactorily) to try to control it *ad hoc* and ignore the remainder.

Meta-analysis theory is clear that positive correlations among data greatly inflate the type I error rates for significance tests of mean effect size (Hedges and Olkin 1985), but we do not estimate mean effect size in our paper, and the sensitivity of our method to statistical nonindependence is probably reduced. In addition, it is likely that the data contain both positive and negative correlations, and we argued in our paper that these diverse correlations may even increase type II error in our analysis (making our statistical test even more conservative, p. 301, Lövei et al., 2009). Finally, if we concede Shelton et al. (2009) their claim that total immature development time and survival are not different between transgenic treatment and control, they must concede either that all of the instar-specific measures are nonsignificant and not significantly correlated or that all significant correlations between instar-specific measures are negative correlations. If any instar-specific measures were strongly positively correlated, the total immature development time or survival would be different between transgenic treatment and control. Any of these concessions contradict the premise of their criticism about nonindependence; thus, their criticism does not logically hold. In short, nonindependence may not be serious for our analysis; indeed, it probably has the opposite effect to that suggested by Shelton et al. (2009), increasing false negatives (type II error) and making our analysis even more conservative.

Given that nonindependence is common in ecological data, an intriguing question remains. If the correlations among the data are unknown, is it better to ignore all potentially correlated data or to extract information from them? It seems that Shelton et al. (2009) prefer the former, when they suggest that the instar-specific survival rates and development times should not even be considered. We prefer to try to extract information from the data because information is costly and hard to come by, and our analysis (Lövei et al. 2009, p. 296) shows that there is additional information in the instar-specific survival rates and development times that is not present in the corresponding summary statistics. This difference in perspective may stem from the competing uses of laboratory data for risk assessment. On the one hand, the data are sometimes used to extrapolate to effects in the field. On the other, they are used to determine whether additional studies are needed to characterize potential effects in the field. The first approach has had dubious success in risk assessment (Suter 2007) and in ecology, which is why we are concerned with the second, more common use of laboratory data in ecological risk assessment. Thus, our published analysis (Lövei et al. 2009) is a first step in characterizing the information associated with laboratory studies and sets the upper bound for what can be inferred from it.

Another useful approach for addressing statistical nonindependence is to determine how much the presumed correlations matter. We are currently approaching this problem by hypothesizing that the data are all highly positively correlated at some level of aggregation (e.g., all instar survival rates are positively correlated) to determine whether the hypothesized correlations affect subsequent statistical inference.

Bitrophic and Tritrophic Analysis

A separation of the responses by direct or indirect exposure, as suggested by Shelton et al. (2009) is useful, because it allows the assessment of host/prey mediated effects separately from direct toxicity. Under bitrophic interaction, the natural enemy directly consumes the toxin. Under tritrophic interaction, the

Table 1. Distribution of responses, from significantly[a] negative to significantly positive, by natural enemies (parasitoids and predators), quantified in all laboratory and greenhouse experiments on the effects of transgenic plants or transgene products on natural enemies

GM product and natural enemy type	Exposure	Number (%) of responses						Random effect		Skewness	
		Negative Significant	Negative not significant	Neutral	Positive not significant	Positive Significant	Total	g^2, 4 *df*	*P*	g^2, 1 *df*	*P*
Cry toxins											
Parasitoids	Bitrophic	3 (6.1)	9 (18.4)	24 (49.0)	13 (26.5)	0 (0.0)	49 (100)	5.2	0.266	0.02	0.888
	Tritrophic	173 (42.5)	102 (25.1)	85 (20.9)	41 (10.1)	6 (1.5)	407 (100)	438.9	0.000	89.4	0.000
Predators	Bitrophic	10 (4.5)	62 (27.7)	98 (43.8)	51 (22.8)	3 (1.3)	224 (100)	32.6	0.000	1.3	0.256
	Tritrophic	28 (8.4)	92 (27.6)	108 (32.4)	91 (27.3)	14 (4.2)	333 (100)	92.8	0.000	0.5	0.479
Proteinase inhibitors											
Parasitoids	Bitrophic	25 (34.2)	19 (26.0)	19 (26.0)	9 (12.3)	1 (1.4)	73 (100)	57.0	0.000	10.8	0.001
	Tritrophic	36 (15.2)	65 (27.4)	71 (30.0)	48 (20.3)	17 (7.2)	237 (100)	91.1	0.000	3.9	0.047
Predators	Bitrophic	5 (22.7)	6 (27.3)	9 (40.9)	2 (9.1)	0 (0.0)	22 (100)	8.3	0.081	2.7	0.104
	Tritrophic	54 (25.7)	37 (17.6)	56 (26.7)	32 (15.2)	31 (14.8)	210 (100)	142.7	0.000	2.6	0.110

Bitrophic and tritrophic experiments were separated. g^2 is the log-linear statistic.
[a] Significance level was set at $P < 0.05$.

natural enemy consumes a prey or host that has directly consumed the toxin. Thus, under tritrophic interaction, there is no direct exposure to the toxin—only indirect exposure. We summarized the responses for bitrophic (direct) and tritrophic (indirect) interactions (Table 1) and found a complex set of responses. The null hypothesis that all effect sizes are 0 was rejected for both bitrophic and tritrophic predator exposure to Cry toxins (Table 1), which contradicts the claims of Romeis et al. (2006b) and Shelton et al. (2009) that nonzero effects only occur in tritrophic studies with predators. For parasitoids, the null hypothesis was rejected for tritrophic but not for bitrophic studies. In addition, the responses of parasitoids in bitrophic experiments were not skewed. Together, these results suggest that parasitoids may be affected by the commercialized Cry toxins only through indirect interaction. For PIs, the null hypothesis was rejected for both bitrophic and tritrophic parasitoid exposure but only for tritrophic predator exposure. The sample size for predator bitrophic studies with PIs is probably too small to support a reliable inference. PIs seem to affect parasitoids by both bi-and tritrophic pathways and predators by at least the tritrophic pathway. Clearly, there is a need for additional analysis of these data.

Summary

We re-emphasize that we have not claimed nor have we implied that laboratory data should be used by themselves to characterize the risks of Cry toxins and PIs to natural enemies. We are concerned that only limited inferences can be drawn by a risk assessment process that relies only on laboratory data to make critical initial decisions about risk. We believe that sound generalizations from the laboratory data concerning the responses to genetically modified plants by natural enemies will emerge as this literature is explored in detail. Just as there are not 10 million different types of population dynamics (Lawton 1992), there are not 10 million different types of responses to GM plants. In our paper, we pointed out that these laboratory data paint a more complex picture than hypothesized by Romeis et al. (2006b) and Shelton et al. (2009). The criticisms of Shelton et al. (2009) of our statistical methods do not stand up to scrutiny and do not invalidate one of our main conclusions: existing Cry toxins and PIs have nonzero effects on natural enemies in the laboratory that need to be understood better. We encourage the reader to critically examine their claims in light of the evidence and explanations given in our original paper (Lövei et al. 2009) and the additional clarifications presented here.

To close on a positive vein, we note several important findings in our paper (Lövei et al. 2009) that were not disputed by Shelton et al. (2009). These include that (1) the data support only limited generalization about the responses of natural enemies to Cry toxins and PIs; (2) there is an overemphasis on five natural enemy species, although the literature is expanding in scope; (3) tests have been conducted in only a few countries; (4) the Cry toxins that have been studied are mainly Cry1Ab, Cry1Ac, and to a lesser extent Cry3Bb, and other commercialized Cry toxins are under-reported. In addition, we found (Lövei et al. 2009) that (5) parasitoids may be more sensitive than predators to the effects of both Cry toxins and PIs, (6) PIs seem to affect natural enemies more than Cry toxins, and (7) Cry toxins and PIs can have beneficial effects on natural enemies.

References Cited

Andow, D. A., G. L. Lövei, G. L., and S. Arpaia. 2006. Ecological risk assessment for *Bt* crops. Nat. Biotechnol. 24: 749–751.

Bai, Y. Y., M. X. Jiang, and J. A. Cheng. 2005. Effects of transgenic *cry1Ab* rice pollen on fitness of *Propylea japonica* (Thunberg). J. Pest Sci. 78: 123–128.

Duan, J. J., M. Marvier, J. Huesing, G. Dively, and Z. Y. Huang. 2008. A meta-analysis of effects of *Bt* crops on honey bees (Hymenoptera: Apidae). PLoS One 3: e1415.

Hedges, L. V., and I. Olkin. 1985. Statistical methods for meta-analysis. Academic, Orlando, FL.

Lawton, J. H. 1992. There are not 10 million kinds of population dynamics. Oikos 63: 337–338.

Lövei, G. L., and S. Arpaia. 2005. The impact of transgenic plants on natural enemies: a critical review of laboratory studies. Entomol. Exp. Appl. 114: 1–14.

Lövei, G. L., D. A. Andow, and S. Arpaia. 2009. Transgenic insecticidal crops and natural enemies: a detailed review of laboratory studies. Environ. Entomol. 38: 293–306.

Marvier, M., C. McCreedy, J. Regetz, and P. Kareiva. 2007. A meta-analysis of effects of *Bt* cotton and maize on nontarget invertebrates. Science 316: 1475–1477.

Naranjo, S. E. 2009. Impact of *Bt* crops on non-target invertebrates and insecticide use patterns. (http://www.cababstractsplus.org/cabreviews/pdfpreload.asp?pdf=E:\MWEB\WWWRoot\Fulltext\htdocs\Uploads\PDF\20093131779.pdf#?pdf=selected.pdf).

Nguyên, V. T., D. A. Andow, G. P. Fitt, E. R. Sujii, E. Fontes, D. R. Just, and L. Q. Quyên. 2008. Challenges and opportunities with GM crops in Vietnam: the case of *Bt* cotton, pp. 1–23. *In* D. A. Andow, A. Hilbeck, and V. T. Nguyên (eds.), Environmental risk assessment of genetically modified organisms: challenges and opportunities with *Bt* cotton in Viet Nam. CABI Publishing, Wallingford, United Kingdom.

O'Callaghan, M., T. R. Glare, E.P.J. Burgess, and L. A. Malone. 2005. Effects of plants genetically modified for insect resistance on nontarget organisms. Annu. Rev. Entomol. 50: 271–292.

Romeis, J., M. Meissle, and F. Bigler. 2006a. Transgenic crops expressing *Bacillus thuringiensis* toxins and biological control. Nat. Biotechnol. 24: 63–71.

Romeis, J., M. Meissle, and F. Bigler. 2006b. Ecological risk assessment for *Bt* crops: Jorg Romeis, Michael Meissle and Franz Bigler respond. Nat. Biotechnol. 24: 751–753.

Shelton, A. M., S. E. Naranjo, J. Romeis, R. L. Hellmich, J. D. Wolt, B. A. Federici, R. Albajes, F. Bigler, E.P.J. Burgess, G. P. Dively, et al. 2009. Setting the record straight: a rebuttal to an erroneous analysis on transgenic insecticidal crops and natural enemies. Transgen. Res. 18: 317–322.

Suter, G. W., III. 2007. Ecological risk assessment, 2nd ed. CRC, Boca Raton, FL.

Wolfenbarger, L. L., S. E. Naranjo, J. G. Lundgren, R. J. Bitzer, and L. S. Watrud. 2008. *Bt* crops effects on functional guilds of non-target arthropods: a meta-analysis. PLoS ONE 3: e2118.

Received 28 May 2009; accepted 3 August 2009.